

**PCT**WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup> :</b> <b>A61N 43/04, C12Q 1/68, C12N 15/00,</b> <b>C07H 21/02, A61K 39/00</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 97/47358</b> <b>(43) International Publication Date:</b> 18 December 1997 (18.12.97)
<b>(21) International Application Number:</b> PCT/US97/09884 <b>(22) International Filing Date:</b> 6 June 1997 (06.06.97)  <b>(30) Priority Data:</b> 60/020,494 11 June 1996 (11.06.96) US 9614731.9 12 July 1996 (12.07.96) GB 60/033,534 20 December 1996 (20.12.96) US  <b>(71) Applicant (for all designated States except US):</b> MERCK & CO., INC. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).  <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> DONNELLY, John, J. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LU, Tong-Ming [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). LIU, Margaret, A. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US). SHIVER, John, W. [US/US]; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).  <b>(74) Common Representative:</b> MERCK & CO., INC.; 126 East Lincoln Avenue, Rahway, NJ 07065 (US).	<b>(81) Designated States:</b> AL, AM, AU, AZ, BA, BB, BG, BR, BY, CA, CN, CU, CZ, EE, GE, HU, IL, IS, JP, KG, KR, KZ, LC, LK, LR, LT, LV, MD, MG, MK, MN, MX, NO, NZ, PL, RO, RU, SG, SI, SK, TJ, TM, TR, TT, UA, US, UZ, VN, YU, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i>	
<b>(54) Title:</b> SYNTHETIC HEPATITIS C GENES		
<b>(57) Abstract</b>  This invention relates to novel methods and formulations of nucleic acid pharmaceutical products, specifically formulations of nucleic acid vaccine products and nucleic acid gene therapy products.		

BEST AVAILABLE COPY

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

**TITLE OF THE INVENTION**  
**SYNTHETIC HEPATITIS C GENES**

**CROSS-REFERENCE TO RELATED APPLICATIONS**

5           Not applicable.

**STATEMENT REGARDING FEDERALLY-SPONSORED R&D**  
Not applicable.

10   **REFERENCE TO MICROFICHE APPENDIX**  
Not applicable.

**FIELD OF THE INVENTION**  
Not applicable.

15   **BACKGROUND OF THE INVENTION**

          This invention relates to novel nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid vaccine products, when introduced directly into muscle cells, induce the  
20   production of immune responses which specifically recognize Hepatitis C virus (HCV).

**Hepatitis C Virus**

          Non-A, Non-B hepatitis (NANBH) is a transmissible disease  
25   (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV), delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three  
30   types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH).

- 2 -

Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of hepatocellular carcinoma. Thus, a need exists for an effective method  
5 for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of Flavivirus particles are known, and are discussed in M. A. Brinton, in  
10 "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid bilayer. Virions are spherical and have a diameter of about 40-50 nm.  
15 Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever virus. They possess positive-stranded RNA genomes (about 11,000  
20 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded,  
25 and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known  
30 viral sequences, small but significant co-linear homologs are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

- 3 -

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the in situ generation of the protein in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

Therefore, this invention contemplates methods for introducing nucleic acids into living tissue to induce expression of proteins. The invention provides a method for introducing viral proteins into the antigen processing pathway to generate virus-specific immune responses including, but not limited to, CTLs. Thus, the need for specific therapeutic agents capable of eliciting desired prophylactic immune responses against viral pathogens is met for HCV virus by this invention. Of particular importance in this therapeutic approach is the ability to induce T-cell immune responses which can prevent infections even of virus strains which are heterologous to the strain from which the antigen gene was obtained. Therefore, this invention provides DNA constructs encoding viral proteins of the hepatitis C virus core, envelope (E1), nonstructural (NS5) genes or any other HCV genes which encode products which generate specific immune responses including but not limited to CTLs.

- 4 -

### DNA Vaccines

- Benvenisty, N., and Reshef, L. [PNAS 83, 9551-9555, (1986)] showed that  $\text{CaCl}_2$ -precipitated DNA introduced into mice intraperitoneally (i.p.), intravenously (i.v.) or intramuscularly (i.m.) could be expressed. The i.m. injection of DNA expression vectors without  $\text{CaCl}_2$  treatment in mice resulted in the uptake of DNA by the muscle cells and expression of the protein encoded by the DNA. The plasmids were maintained episomally and did not replicate. Subsequently, persistent expression has been observed after i.m. injection in skeletal muscle of rats, fish and primates, and cardiac muscle of rats. The technique of using nucleic acids as therapeutic agents was reported in WO90/11092 (4 October 1990), in which polynucleotides were used to vaccinate vertebrates.
- It is not necessary for the success of the method that immunization be intramuscular. The introduction of gold microprojectiles coated with DNA encoding bovine growth hormone (BGH) into the skin of mice resulted in production of anti-BGH antibodies in the mice. A jet injector has been used to transfect skin, muscle, fat, and mammary tissues of living animals. Various methods for introducing nucleic acids have been reviewed. Intravenous injection of a DNA:cationic liposome complex in mice was shown by Zhu et al., [Science 261:209-211 (9 July 1993)] to result in systemic expression of a cloned transgene. Ulmer et al., [Science 259:1745-1749, (1993)] reported on the heterologous protection against influenza virus infection by intramuscular injection of DNA encoding influenza virus proteins.

The need for specific therapeutic and prophylactic agents capable of eliciting desired immune responses against pathogens and tumor antigens is met by the instant invention. Of particular importance in this therapeutic approach is the ability to induce T-cell immune responses which can prevent infections or disease caused even by virus strains which are heterologous to the strain from which the antigen gene was obtained. This is of particular concern when dealing with HIV as this virus has been recognized to mutate rapidly and many virulent isolates have been identified [see, for example, LaRosa et al.,

- 5 -

Science 249:932-935 (1990), identifying 245 separate HIV isolates]. In response to this recognized diversity, researchers have attempted to generate CTLs based on peptide immunization. Thus, Takahashi et al., [Science 255:333-336 (1992)] reported on the induction of broadly  
5 cross-reactive cytotoxic T cells recognizing an HIV envelope (gp160) determinant. However, those workers recognized the difficulty in achieving a truly cross-reactive CTL response and suggested that there is a dichotomy between the priming or restimulation of T cells, which is very stringent, and the elicitation of effector function, including  
10 cytotoxicity, from already stimulated CTLs.

Wang et al. reported on elicitation of immune responses in mice against HIV by intramuscular inoculation with a cloned, genomic (unspliced) HIV gene. However, the level of immune responses achieved in these studies was very low. In addition, the Wang et al.,  
15 DNA construct utilized an essentially genomic piece of HIV encoding contiguous Tat/REV-gp160-Tat/REV coding sequences. As is described in detail below, this is a suboptimal system for obtaining high-level expression of the gp160. It also is potentially dangerous because expression of Tat contributes to the progression of Kaposi's Sarcoma.  
20

WO 93/17706 describes a method for vaccinating an animal against a virus, wherein carrier particles were coated with a gene construct and the coated particles are accelerated into cells of an animal.

The instant invention contemplates any of the known methods for introducing polynucleotides into living tissue to induce  
25 expression of proteins. However, this invention provides a novel immunogen for introducing proteins into the antigen processing pathway to efficiently generate specific CTLs and antibodies.

#### Codon Usage and Codon Context

30 The codon pairings of organisms are highly nonrandom, and differ from organism to organism. This information is used to construct and express altered or synthetic genes having desired levels of translational efficiency, to determine which regions in a genome are protein coding regions, to introduce translational pause sites into

- 6 -

heterologous genes, and to ascertain relationship or ancestral origin of nucleotide sequences

The expression of foreign heterologous genes in transformed organisms is now commonplace. A large number of mammalian genes, including, for example, murine and human genes, have been successfully inserted into single celled organisms. Standard techniques in this regard include introduction of the foreign gene to be expressed into a vector such as a plasmid or a phage and utilizing that vector to insert the gene into an organism. The native promoters for such genes are commonly replaced with strong promoters compatible with the host into which the gene is inserted. Protein sequencing machinery permits elucidation of the amino acid sequences of even minute quantities of native protein. From these amino acid sequences, DNA sequences coding for those proteins can be inferred. DNA synthesis is also a rapidly developing art, and synthetic genes corresponding to those inferred DNA sequences can be readily constructed.

Despite the burgeoning knowledge of expression systems and recombinant DNA, significant obstacles remain when one attempts to express a foreign or synthetic gene in an organism. Many native, active proteins, for example, are glycosylated in a manner different from that which occurs when they are expressed in a foreign host. For this reason, eukaryotic hosts such as yeast may be preferred to bacterial hosts for expressing many mammalian genes. The glycosylation problem is the subject of continuing research.

Another problem is more poorly understood. Often translation of a synthetic gene, even when coupled with a strong promoter, proceeds much less efficiently than would be expected. The same is frequently true of exogenous genes foreign to the expression organism. Even when the gene is transcribed in a sufficiently efficient manner that recoverable quantities of the translation product are produced, the protein is often inactive or otherwise different in properties from the native protein.

- 7 -

It is recognized that the latter problem is commonly due to differences in protein folding in various organisms. The solution to this problem has been elusive, and the mechanisms controlling protein folding are poorly understood.

5       The problems related to translational efficiency are believed to be related to codon context effects. The protein coding regions of genes in all organisms are subject to a wide variety of functional constraints, some of which depend on the requirement for encoding a properly functioning protein, as well as appropriate  
10 translational start and stop signals. However, several features of protein coding regions have been discerned which are not readily understood in terms of these constraints. Two important classes of such features are those involving codon usage and codon context.

It is known that codon utilization is highly biased and varies  
15 considerably between different organisms. Codon usage patterns have been shown to be related to the relative abundance of tRNA isoacceptors. Genes encoding proteins of high versus low abundance show differences in their codon preferences. The possibility that biases in codon usage alter peptide elongation rates has been widely discussed.  
20 While differences in codon use are associated with differences in translation rates, direct effects of codon choice on translation have been difficult to demonstrate. Other proposed constraints on codon usage patterns include maximizing the fidelity of translation and optimizing the kinetic efficiency of protein synthesis.

25       Apart from the non-random use of codons, considerable evidence has accumulated that codon/anticodon recognition is influenced by sequences outside the codon itself, a phenomenon termed "codon context." There exists a strong influence of nearby nucleotides on the efficiency of suppression of nonsense codons as well as missense codons.  
30 Clearly, the abundance of suppressor activity in natural bacterial populations, as well as the use of "termination" codons to encode selenocysteine and phosphoserine require that termination be context-dependent. Similar context effects have been shown to influence the fidelity of translation, as well as the efficiency of translation initiation.

- 8 -

Statistical analyses of protein coding regions of E. coli have demonstrate another manifestation of "codon context." The presence of a particular codon at one position strongly influences the frequency of occurrence of certain nucleotides in neighboring codons, and these  
5 context constraints differ markedly for genes expressed at high versus low levels. Although the context effect has been recognized, the predictive value of the statistical rules relating to preferred nucleotides adjacent to codons is relatively low. This has limited the utility of such nucleotide preference data for selecting codons to effect desired levels  
10 of translational efficiency.

The advent of automated nucleotide sequencing equipment has made available large quantities of sequence data for a wide variety of organisms. Understanding those data presents substantial difficulties. For example, it is important to identify the coding regions of the  
15 genome in order to relate the genetic sequence data to protein sequences. In addition, the ancestry of the genome of certain organisms is of substantial interest. It is known that genomes of some organisms are of mixed ancestry. Some sequences that are viral in origin are now stably incorporated into the genome of eukaryotic organisms. The viral  
20 sequences themselves may have originated in another substantially unrelated species. An understanding of the ancestry of a gene can be important in drawing proper analogies between related genes and their translation products in other organisms.

There is a need for a better understanding of codon context  
25 effects on translation, and for a method for determining the appropriate codons for any desired translational effect. There is also a need for a method for identifying coding regions of the genome from nucleotide sequence data. There is also a need for a method for controlling protein folding and for insuring that a foreign gene will fold appropriately  
30 when expressed in a host. Genes altered or constructed in accordance with desired translational efficiencies would be of significant worth.

Another aspect of the practice of recombinant DNA techniques for the expression by microorganisms of proteins of industrial and pharmaceutical interest is the phenomenon of "codon

- 9 -

preference". While it was earlier noted that the existing machinery for gene expression in genetically transformed host cells will "operate" to construct a given desired product, levels of expression attained in a microorganism can be subject to wide variation, depending in part on specific alternative forms of the amino acid-specifying genetic code present in an inserted exogenous gene. A "triplet" codon of four possible nucleotide bases can exist in 64 variant forms. That these forms provide the message for only 20 different amino acids (as well as transcription initiation and termination) means that some amino acids can be coded for by more than one codon. Indeed, some amino acids have as many as six "redundant", alternative codons while some others have a single, required codon. For reasons not completely understood, alternative codons are not at all uniformly present in the endogenous DNA of differing types of cells and there appears to exist a variable natural hierarchy or "preference" for certain codons in certain types of cells.

As one example, the amino acid leucine is specified by any of six DNA codons including CTA, CTC, CTG, CTT, TTA, and TTG (which correspond, respectively, to the mRNA codons, CUA, CUC, CUG, CUU, UUA and UUG). Exhaustive analysis of genome codon frequencies for microorganisms has revealed endogenous DNA of E. coli most commonly contains the CTG leucine-specifying codon, while the DNA of yeasts and slime molds most commonly includes a TTA leucine-specifying codon. In view of this hierarchy, it is generally held that the likelihood of obtaining high levels of expression of a leucine-rich polypeptide by an E. coli host will depend to some extent on the frequency of codon use. For example, a gene rich in TTA codons will in all probability be poorly expressed in E. coli, whereas a CTG rich gene will probably highly express the polypeptide. Similarly, when yeast cells are the projected transformation host cells for expression of a leucine-rich polypeptide, a preferred codon for use in an inserted DNA would be TTA.

The implications of codon preference phenomena on recombinant DNA techniques are manifest, and the phenomenon may

- 10 -

serve to explain many prior failures to achieve high expression levels of exogenous genes in successfully transformed host organisms-a less "preferred" codon may be repeatedly present in the inserted gene and the host cell machinery for expression may not operate as efficiently.

- 5 This phenomenon suggests that synthetic genes which have been designed to include a projected host cell's preferred codons provide a preferred form of foreign genetic material for practice of recombinant DNA techniques.

10 Protein Trafficking

- The diversity of function that typifies eukaryotic cells depends upon the structural differentiation of their membrane boundaries. To generate and maintain these structures, proteins must be transported from their site of synthesis in the endoplasmic reticulum to
- 15 predetermined destinations throughout the cell. This requires that the trafficking proteins display sorting signals that are recognized by the molecular machinery responsible for route selection located at the access points to the main trafficking pathways. Sorting decisions for most proteins need to be made only once as they traverse their
- 20 biosynthetic pathways since their final destination, the cellular location at which they perform their function, becomes their permanent residence.

- Maintenance of intracellular integrity depends in part on the selective sorting and accurate transport of proteins to their correct destinations. Over the past few years the dissection of the molecular
- 25 machinery for targeting and localization of proteins has been studied vigorously. Defined sequence motifs have been identified on proteins which can act as 'address labels'. A number of sorting signals have been found associated with the cytoplasmic domains of membrane proteins.

30

**SUMMARY OF THE INVENTION**

This invention relates to novel formulations of nucleic acid pharmaceutical products, specifically nucleic acid vaccine products. The nucleic acid products, when introduced directly into muscle cells,

- 11 -

induce the production of immune responses which specifically recognize Hepatitis C virus (HCV).

#### BRIEF DESCRIPTION OF THE DRAWINGS

- 5           Figure 1 shows the nucleotide sequence of the VIRa vector.  
          Figure 2 is a diagram of the VIRa vector.  
          Figure 3 is a diagram of the Vtpa vector.  
          Figure 4 is the VUb vector  
          Figure 5 shows an optimized sequence of the HCV core  
10   antigen.  
          Figure 6 shows VIRa.HCV1CorePAb, Vtpa.HCV1CorePAb  
          and VUb.HCV1CorePAb.  
          Figure 7 shows the Hepatitis C Virus Core Antigen  
          Sequence.  
15           Figure 8 shows codon utilization in human protein-coding  
          sequences (from Lathe et al.).  
          Figure 9 shows an optimized sequence of the HCV E1  
          protein.  
          Figure 10 shows an optimized sequence of the HCV E2  
20   protein.  
          Figure 11 shows an optimized sequence of the HCV E1 +E2  
          proteins.  
          Figure 12 shows an optimized sequence of the HCV NS5a  
          protein.  
25           Figure 13 shows an optimized sequence of the HCV NS5b  
          protein.

#### DETAILED DESCRIPTION OF THE INVENTION

- 30           This invention relates to novel formulations of nucleic acid  
          pharmaceutical products, specifically nucleic acid vaccine products.  
          The nucleic acid vaccine products, when introduced directly into muscle  
          cells, induce the production of immune responses which specifically  
          recognize Hepatitis C virus (HCV).

- 12 -

Non-A, Non-B hepatitis (NANBH) is a transmissible disease (or family of diseases) that is believed to be virally induced, and is distinguishable from other forms of virus-associated liver disease, such as those caused by hepatitis A virus (HAV), hepatitis B virus (HBV),  
5 delta hepatitis virus (HDV), cytomegalovirus (CMV) or Epstein-Barr virus (EBV). Epidemiologic evidence suggests that there may be three types of NANBH: the water-borne epidemic type; the blood or needle associated type; and the sporadically occurring (community acquired) type. However, the number of causative agents is unknown. Recently, a  
10 new viral species, hepatitis C virus (HCV) has been identified as the primary (if not only) cause of blood-associated NANBH (BB-NANBH). Hepatitis C appears to be the major form of transfusion-associated hepatitis in a number of countries, including the United States and Japan. There is also evidence implicating HCV in induction of  
15 hepatocellular carcinoma. Thus, a need exists for an effective method for preventing or treating HCV infection: currently, there is none.

The HCV may be distantly related to the flaviviridae. The Flavivirus family contains a large number of viruses which are small, enveloped pathogens of man. The morphology and composition of  
20 Flavivirus particles are known, and are discussed in M. A. Brinton, in "The Viruses: The Togaviridae And Flaviviridae" (Series eds. Fraenkel-Conrat and Wagner, vol. eds. Schlesinger and Schlesinger, Plenum Press, 1986), pp. 327-374. Generally, with respect to morphology, Flaviviruses contain a central nucleocapsid surrounded by a lipid  
25 bilayer. Virions are spherical and have a diameter of about 40-50 nm. Their cores are about 25-30 nm in diameter. Along the outer surface of the virion envelope are projections measuring about 5-10 nm in length with terminal knobs about 2 nm in diameter. Typical examples of the family include Yellow Fever virus, West Nile virus, and Dengue Fever  
30 virus. They possess positive-stranded RNA genomes (about 11,000 nucleotides) that are slightly larger than that of HCV and encode a polyprotein precursor of about 3500 amino acids. Individual viral proteins are cleaved from this precursor polypeptide.

- 13 -

The genome of HCV appears to be single-stranded RNA containing about 10,000 nucleotides. The genome is positive-stranded, and possesses a continuous translational open reading frame (ORF) that encodes a polyprotein of about 3,000 amino acids. In the ORF, the structural proteins appear to be encoded in approximately the first quarter of the N-terminal region, with the majority of the polyprotein attributed to non-structural proteins. When compared with all known viral sequences, small but significant co-linear homologics are observed with the nonstructural proteins of the Flavivirus family, and with the pestiviruses (which are now also considered to be part of the Flavivirus family).

Intramuscular inoculation of polynucleotide constructs, i.e., DNA plasmids encoding proteins have been shown to result in the generation of the encoded protein in situ in muscle cells. By using cDNA plasmids encoding viral proteins, both antibody and CTL responses were generated, providing homologous and heterologous protection against subsequent challenge with either the homologous or cross-strain protection, respectively. Each of these types of immune responses offers a potential advantage over existing vaccination strategies. The use of PNVs (polynucleotide vaccines) to generate antibodies may result in an increased duration of the antibody responses as well as the provision of an antigen that can have both the exact sequence of the clinically circulating strain of virus as well as the proper post-translational modifications and conformation of the native protein (vs. a recombinant protein). The generation of CTL responses by this means offers the benefits of cross-strain protection without the use of a live potentially pathogenic vector or attenuated virus.

The standard techniques of molecular biology for preparing and purifying DNA constructs enable the preparation of the DNA therapeutics of this invention. While standard techniques of molecular biology are therefore sufficient for the production of the products of this invention, the specific constructs disclosed herein provide novel therapeutics which surprisingly produce cross-strain

- 14 -

protection, a result heretofore unattainable with standard inactivated whole virus or subunit protein vaccines.

The amount of expressible DNA to be introduced to a vaccine recipient will depend on the strength of the transcriptional and translational promoters used in the DNA construct, and on the immunogenicity of the expressed gene product. In general, an immunologically or prophylactically effective dose of about 1  $\mu$ g to 1 mg, and preferably about 10  $\mu$ g to 300  $\mu$ g is administered directly into muscle tissue. Subcutaneous injection, intradermal introduction, impression through the skin, and other modes of administration such as intraperitoneal, intravenous, or inhalation delivery are also contemplated. It is also contemplated that booster vaccinations are to be provided.

The DNA may be naked, that is, unassociated with any proteins, adjuvants or other agents which impact on the recipient's immune system. In this case, it is desirable for the DNA to be in a physiologically acceptable solution, such as, but not limited to, sterile saline or sterile buffered saline. Alternatively, the DNA may be associated with surfactants, liposomes, such as lecithin liposomes or other liposomes known in the art, as a DNA-liposome mixture, (see for example WO93/24640) or the DNA may be associated with an adjuvant known in the art to boost immune responses, such as a protein or other carrier. Agents which assist in the cellular uptake of DNA, such as, but not limited to, calcium ions, detergents, viral proteins and other transfection facilitating agents may also be used to advantage. These agents are generally referred to as transfection facilitating agents and as pharmaceutically acceptable carriers. As used herein, the term gene refers to a segment of nucleic acid which encodes a discrete polypeptide. The term pharmaceutical, and vaccine are used interchangeably to indicate compositions useful for inducing immune responses. The terms construct, and plasmid are used interchangeably. The term vector is used to indicate a DNA into which genes may be cloned for use according to the method of this invention.

- 15 -

The following examples are provided to further define the invention, without limiting the invention to the specifics of the examples.

5

**EXAMPLE 1****V1J EXPRESSION VECTORS:**

V1J is derived from vectors V1 and pUC18, a commercially available plasmid. V1 was digested with SspI and EcoRI restriction enzymes producing two fragments of DNA. The smaller of these fragments, containing the CMVintA promoter and Bovine Growth Hormone (BGH) transcription termination elements which control the expression of heterologous genes, was purified from an agarose electrophoresis gel. The ends of this DNA fragment were then "blunted" using the T4 DNA polymerase enzyme in order to facilitate its ligation to another "blunt-ended" DNA fragment.

pUC18 was chosen to provide the "backbone" of the expression vector. It is known to produce high yields of plasmid, is well-characterized by sequence and function, and is of minimum size. We removed the entire *lac* operon from this vector, which was unnecessary for our purposes and may be detrimental to plasmid yields and heterologous gene expression, by partial digestion with the HaeII restriction enzyme. The remaining plasmid was purified from an agarose electrophoresis gel, blunt-ended with the T4 DNA polymerase, treated with calf intestinal alkaline phosphatase, and ligated to the CMVintA/BGH element described above. Plasmids exhibiting either of two possible orientations of the promoter elements within the pUC backbone were obtained. One of these plasmids gave much higher yields of DNA in *E. coli* and was designated V1J. This vector's structure was verified by sequence analysis of the junction regions and was subsequently demonstrated to give comparable or higher expression of heterologous genes compared with V1. The ampicillin resistance marker was replaced with the neomycin resistance marker to yield vector V1Jneo.

- 16 -

An Sfi I site was added to V1Jneo to facilitate integration studies. A commercially available 13 base pair Sfi I linker (New England BioLabs) was added at the Kpn I site within the BGH sequence of the vector. V1Jneo was linearized with Kpn I, gel purified, blunted by T4 DNA polymerase, and ligated to the blunt Sfi I linker. Clonal isolates were chosen by restriction mapping and verified by sequencing through the linker. The new vector was designated V1Jns. Expression of heterologous genes in V1Jns (with Sfi I) was comparable to expression of the same genes in V1Jneo (with Kpn I).

Vector V1Ra (Sequence is shown in Figure 1; map is shown in Figure 2) was derived from vector V1R, a derivative of the V1Jns vector. Multiple cloning sites (*Bgl*II, *Kpn*I, *Eco*RV, *Eco*RI, *Sal*I, and *Not*I) were introduced into V1R to create the V1Ra vector to improve the convenience of subcloning. V1Ra vector derivatives containing the tpa leader sequence and ubiquitin sequence were generated (Vtpa (Figure 3) and Vub (Figure 4), respectively). Expression of viral antigen from Vtpa vector will target the antigen protein into the exocytic pathway, thus producing a secretable form of the antigen proteins. These secreted proteins are likely to be captured by professional antigen presenting cells, such as macrophages and dendritic cells, and processed and presented by class II molecules to activate CD4<sup>+</sup> Th cells. They also are more likely to efficiently simulate antibody responses. Expression of viral antigen through VUb vector will produce a ubiquitin and antigen fusion protein. The uncleavable ubiquitin segment (glycine to alanine change at the cleavage site, Butt et al., JBC 263:16364, 1988) will target the viral antigen to ubiquitin-associated proteasomes for rapid degradation. The resulting peptide fragments will be transported into the ER for antigen presentation by class I molecules. This modification is attempted to enhance the class I molecule-restricted CTL responses against the viral antigen (Townsend et al, JEM 168:1211, 1988).

- 17 -

**EXAMPLE 2**  
**DESIGN AND CONSTRUCTION OF THE SYNTHETIC GENES**

**A. Design of Synthetic Gene Segments for HCV Gene Expression:**

- 5           Gene segments were converted to sequences having identical translated sequences (except where noted) but with alternative codon usage as defined by R. Lathe in a research article from *J. Molec. Biol.* Vol. 183, pp. 1-12 (1985) entitled "Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data: Theoretical and
- 10   Practical Considerations". The methodology described below was based on our hypothesis that the known inability to express a gene efficiently in mammalian cells is a consequence of the overall transcript composition. Thus, using alternative codons encoding the same protein sequence may remove the constraints on HCV gene expression.
- 15   Inspection of the codon usage within HCV genome revealed that a high percentage of codons were among those infrequently used by highly expressed human genes. The specific codon replacement method employed may be described as follows employing data from Lathe et al.:
- 20           1.    Identify placement of codons for proper open reading frame.
2.    Compare wild type codon for observed frequency of use by human genes (refer to Table 3 in Lathe et al.).
3.    If codon is not the most commonly employed,
- 25   replace it with an optimal codon for high expression based on data in Table 5.
4.    Inspect the third nucleotide of the new codon and the first nucleotide of the adjacent codon immediately 3'- of the first. If a 5'-CG-3' pairing has been created by the new codon selection, replace it
- 30   with the choice indicated in Table 5.
5.    Repeat this procedure until the entire gene segment has been replaced.
6.    Inspect new gene sequence for undesired sequences generated by these codon replacements (e.g., "ATTTA" sequences,

- 18 -

inadvertent creation of intron splice recognition sites, unwanted restriction enzyme sites, etc.) and substitute codons that eliminate these sequences.

7. Assemble synthetic gene segments and test for improved expression.

#### B. HCV CORE ANTIGEN SEQUENCE

The consensus core sequence of HCV was adopted from a generalized core sequence reported by Bukh et al. (PNAS, 91:8239, 1994). This core sequence contains all the identified CTL epitopes in both human and mouse. The gene is composed of 573 nucleotides and encodes 191 amino acids. The predicted molecular weight is about 23 kDa.

- 15 The codon replacement was conducted to eliminate codons which may hinder the expression of the HCV core protein in transfected mammalian cells in order to maximize the translational efficiency of DNA vaccine. Twenty three point two percent (23.2%) of nucleotide sequence (133 out of 573 nucleotides) were altered, resulting in changes of 61.3% of the codons (117 out 191 codons) in the core antigen sequence. The optimized nucleotide sequence of HCV core is shown in Figure 5.

#### C. CONSTRUCTION OF THE SYNTHETIC CORE GENE

- 25 The optimized HCV core gene (Figure 5) was constructed as a synthetic gene annealed from multiple synthetic oligonucleotides. To facilitate the identification and evaluation of the synthetic gene expression in cell culture and its immunogenicity in mice, a CTL epitope derived from influenza virus nucleoprotein residues 366-374 and an antibody epitope sequence derived from SV40 T antigen residues 684-698 were tagged to the carboxyl terminal of the core sequence (Figure 6). For clinical use it may be desired to express the core sequence without the nucleoprotein 366-374 and SV40 T 684-698 sequences. For this reason, the sequence of the two epitopes is flanked by two *EcoRI* sites which will be used to excise this fragment of

- 19 -

sequence at a later time. Thus an embodiment of the invention for clinical use could consist of the VIRa.HCV1CorePAb, Vtpa.HCV1CorePAb, or VUb.HCV1CorePAb plasmids that had been cut with EcoRI, annealed, and ligated to yield plasmids

5 VIRa.HCV1Core, Vtpa.HCV1Core, and VUb.HCV1Core.

The synthetic gene was built as three separate segments in three vectors, nucleotides 1 to 80 in VIRa, nucleotides 80 to 347 (*Bst*XI site) in pUC18, and nucleotides 347 to 573 plus the two epitope sequence in pUC18. All the segments were verified by DNA  
10 sequencing, and joined together in VIRa vector.

#### D. HCV Gene Expression Constructs:

In each case, the junction sequences from the 5' promoter region (CMVintA) into the cloned gene is shown. The position at which  
15 the junction occurs is demarcated by a "/", which does not represent any discontinuity in the sequence.

The nomenclature for these constructs follows the convention: "Vector name-HCV strain-gene".

20

#### VIRa.HCV1.CorePAb

---IntA--AGA TCT ACC / ATG AGC--HCV.Core--GCC / GAA TTC GCT TCC--  
PAb Sequence--TAA / ACC CGG GAA TTC TAA A / GTC GAC--BGH---

25

#### Vtpa.HCV1.CorePAb

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH---

30

#### VUb.HCV1.CorePAb.

---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG / ATG AGC--  
HCV.Core--GCC / GAA TTC GCT TCC--PAb Sequence--TAA / ACC CGG GAA  
TTC TAA A / GTC GAC--BGH---

- 20 -

**VIRa.HCV1.Core**

---IntA--AGA TCT ACC / ATG AGC--HCV.Core--GCC / TAA A / GTC GAC--  
BGH---

5

**Vtpa.HCV1.Core**

---IntA--ATC ACC / ATG GAT--tpa leader--GAG ATC-TTC / ATG AGC--  
HCV.Core--GCC / TAA A / GTC GAC--BGH---

10

**VUb.HCV1.Core**

---IntA--AGA TCC ACC / ATG CAG--Ubiquitin--GGT GCA GAT CTG/ ATG AGC--  
HCV.Core--GCC / TAA A / GTC GAC--BGH---

**E. OTHER SYNTHETIC HCV GENES**

- 15      Using similar codon optimization techniques, synthetic  
genes encoding the HCV E1 (Figure 9), HCV E2 (Figure 10), HCV  
E1+E2 (Figure 11), HCV NS5a (Figure 12) and HCV NS5b (Figure 13)  
proteins were created.

## WHAT IS CLAIMED:

1. A synthetic polynucleotide comprising a DNA sequence encoding an HCV protein selected from the group consisting of HCV core protein, HCV E1 protein, HCV E1+E2 protein, HCV NS5a protein, HCV NS5b protein and fragments thereof, the DNA sequence comprising codons optimized for expression in a vertebrate host.
2. A plasmid vector comprising the polynucleotide of Claim 1, the plasmid vector being suitable for immunization of a vertebrate host.
3. The polynucleotide of Claim 1 which is HCV genotype I/1a core.

## 4. The polynucleotide of Claim 1 having the sequence

```

1 ATGAGTACCA ACCCAAGTTC CAGATTAAG ACCCAAGTTC ACACCAAGTC GAGGTCATC CATCTAAAT TCTCTAGG 80
81 AGGCTACATC GTTGAAGAGG TTACCTACT CCAAGTAAAT GGTCTAAGC TGGAGTAAAG GGTACCGG AAGACCTTTC 160
161 AAGGCTTACA GTCCAGTGGC AGTACCTAC CCAATCTTAA GCTCTCTGCT CCGGAGTAA CAGCTCTTCT CAGCTCTTCT 240
241 TACCTCTGCT CCAATCTTAA CCAATCTTAA TTTGCTCTG CCGCTCTTCT GCTCTCTTCT GAGGCTTCTG GAGGCTTCTG 320
321 GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 400
401 TGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 480
481 GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 560
561 GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 571

```

## 5. The plasmid vector of Claim 2 having the sequence

```

1 CATATTGCT ATTCTTATT GATATCTTC TATCATATA ATAATATTA CATTATATT GCTCATCTC CAATATTAC 80
81 GCTATCTTA CATATATT TACTATTTA TATATTTAA TCAATTACG GCTCATATT TCAATATTA TATATTTAA 160
161 TCCCTGTTAC ATAACTTAC TCAATTTTC GCTCTCTGCT CCAATCTTAA GCTCTCTGCT CCAATCTTAA TCAATTTTC 240
241 TATCTTCCA TATTAATTC AATATTTCT TTTGCTCTG CCAATCTTAA GCTCTCTGCT CCAATCTTAA TCAATTTTC 320
321 ATACATCAAT CTATATTTA TCCCAATTC GCTCTCTGCT CCAATCTTAA GCTCTCTGCT CCAATCTTAA TCAATTTTC 400
401 ATACATCAAT CTATATTTA TCCCAATTC GCTCTCTGCT CCAATCTTAA GCTCTCTGCT CCAATCTTAA TCAATTTTC 480
481 TGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 560
561 GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG GAGGCTTCTG 571

```

- 22 -

2481 GTATTCTTAA 3322222222 GGGGAGGAGA GGAAGGAGGA GATTTGAGAA GAGTAATGGA GGGATGTTG GATGAGTTG 2560  
 2482 GATTTTATG 3322222222 AGGAGGTTA ATTAAAGGAG GAGGAGGTTT AGGAGGTTG TGAAGAAATG AGGAGGTTT 2561  
 2483 TATTAATGTA AAAAGGAGG 3322222222 TTTTTCATA GAGGAGGTTT GGGGAGGAG GATTAAGAAA ATTGAATGTT 2562  
 2721 AAGTACAGAT 3322222222 GAGAGGAGT ATAAAGATAG GAGGAGGTTT GGGGAGGAG GGGGAGGTTT 2580  
 2801 TTTTACATTT GGGGAGGAG GATACGTTT GGGGAGGTTT TTTTTCATA AGGAGGTTT TTTTACATTT 2581  
 2881 AATATATTA GTTTCATTA GGGGAGGTTT TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2582  
 2961 GTTATGAGT AACTATGTT TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2583  
 3041 TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2584  
 3121 TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2585  
 3201 TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2586  
 3281 GTTATGAGT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2587  
 3361 AAAATGCAAT TTATTCATAT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2588  
 3441 GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2589  
 3521 TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2590  
 3601 AATATATTT TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2591  
 3681 TTTTACATTT TTTTACATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2592  
 3761 GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2593  
 3841 GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2594  
 3921 GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2595  
 4001 GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2596  
 4081 TAAATATTA TTTTACATTT AATTAATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2597  
 4161 TATTAATTT TATTAATTT GAGAGTTT TTTTACATTT TTTTACATTT TTTTACATTT TTTTACATTT 2598  
 4241 TATTAATTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT GGGGAGGTTT 2599

25 6. The polynucleotide of Claim 4 from which the PAB sequence has been removed.

30 7. The plasmid vector of Claim 5 from which the PAB sequence has been removed.

35 8. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1 ng and 100 mg of the polynucleotide of Claim 1 into the tissue of the vertebrate.

40 9. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into the tissue of a vertebrate the polynucleotide of Claim 1.

45 10. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 1 and a pharmaceutically acceptable carrier.

11. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 1 into the tissue of said primate and concurrently administering interleukin-12 parenterally.

- 23 -

12. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation and effector functions including lymphokine secretion specific to HCV antigens which  
5 comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 1.

13. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of  
10 Claim 1 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..

14. A pharmaceutical composition comprising the polynucleotide of Claim 1.  
15

15. A method of inducing an immune response comprising administering the polynucleotide of Claim 1 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant,  
20 recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.

16. A method for inducing immune responses in a vertebrate against HCV epitopes which comprises introducing between 1  
25 ng and 100 mg of the polynucleotide of Claim 2 into the tissue of the vertebrate.

17. A method for inducing immune responses against infection or disease caused by HCV which comprises introducing into  
30 the tissue of a vertebrate the polynucleotide of Claim 2.

18. A vaccine for inducing immune responses against HCV infection which comprises the polynucleotide of Claim 2 and a pharmaceutically acceptable carrier.

- 24 -

19. A method for inducing anti-HCV immune responses in a primate which comprises introducing the polynucleotide of Claim 2 into the tissue of said primate and concurrently administering interleukin 12 parenterally.
20. A method of inducing an antigen presenting cell to stimulate cytotoxic and helper T-cell proliferation and effector functions including lymphokine secretion specific to HCV antigens which comprises exposing cells of a vertebrate in vivo to the polynucleotide of Claim 2.
21. A method of treating a patient in need of such treatment comprising administering to the patient the polynucleotide of Claim 2 in combination with interferon-alpha, Ribavirin, Zidovudine, or other pharmaceutically acceptable antiviral agents..
22. A pharmaceutical composition comprising the polynucleotide of Claim 2.
23. A method of inducing an immune response comprising administering the polynucleotide of Claim 2 to a patient, the administration of the polynucleotide antedating or coinciding or following administration to the patient of a subunit, recombinant, recombinant live vector, inactivated, recombinant inactivated vector, or live attenuated HCV vaccine.
24. The vector of Claim 2 which is selected from VIRa.HCV1CorePAb, Vtpa.HCV1CorePAb, VUb.HCV1CorePAb, VIRa.HCV1Core, Vtpa.HCV1Core and VUb.HCV1Core.
25. A pharmaceutical composition comprising the vector of Claim 21.

- 25 -

26. The DNA sequence of Claim 1 selected from the group consisting of a nucleotide sequence shown in Figure 5, Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13.

1/22

1	GATATTGGCT	10	ATGGCCAT	20	GCATACGTTG	30	TATCCATATC	40	ATAATATGTA	50	CATTATATT	60	GGCTCATGTC	70	CAACATTACC	80
81	GCCATGTTGA	10	CATTGATTAT	20	TGACTAGTTA	30	TAAATAGTAA	40	TCAATTACGG	50	GGTCATTAGT	60	TCATAGCCCA	70	TATATGGAGT	80
161	TCCGCGTTAC	10	ATAACTTAGC	20	GTAATGGCC	30	CGCTGGCTG	40	ACGCCCAAC	50	GACCCCGCC	60	CATTGACGTC	70	AATAATGACG	80
241	TATGTTCCCA	10	TAGTAACGCC	20	AATAGGGACT	30	TTCCATTGAC	40	GTAATGGGT	50	GGAGTATTTA	60	CGGTAACACTG	70	CCCACCTGGC	80
321	AGTACATCAA	10	GTGTATCATA	20	TGCCAAGTAC	30	GCCCCCTATT	40	GACGTCAATG	50	ACGGTAAATG	60	GCCCGCCTGG	70	CATTATGCCC	80
401	AGTACATGAC	10	CTTATGGGAC	20	TTTCCTACTT	30	GGCAGTACAT	40	CTAGGTATTA	50	GTCTATGCTA	60	TTACCATGGT	70	GATCGGTTT	80
481	TGGCAGTACA	10	TCAATGGGGC	20	TGGATAGCGG	30	TTTGACTCAC	40	GGGGATTTCC	50	AAGTCTCCAC	60	CCCATTGACG	70	TCAATGGGAG	80
561	TTTGTTTGG	10	CACCAAAATC	20	AACGGGACTT	30	TCCAATAATG	40	CGTAACAAT	50	CGCCCCCAT	60	GACGCCAATG	70	GGCGGTAGGC	80
641	GTGTACGGTG	10	GGAGGTCTAT	20	ATAAGCAGAG	30	CTCGTTTAGT	40	GAACCGTCAG	50	ATCGCCTGGA	60	GACGCCATCC	70	ACGCTGTTTT	80
721	GACCTCCATA	10	GAAGACACCG	20	GGACCGATCC	30	AGCCTCCGG	40	GCCGGGAACG	50	GTGCATTGGA	60	ACGCGGATTC	70	CCCGTGCCAA	80
801	GAGTGACGTA	10	AGTACCGCT	20	ATAGAGTCTA	30	TAGGCCACCC	40	CCCTTGGCTT	50	CTTATGCAIG	60	CTATACTGTT	70	TTTGGCTTGG	80
881	GGTCTATACA	10	CCCCGGCTTC	20	CTCATGTTAT	30	AGGTGATGGT	40	ATAGCTTAGC	50	CTATAGGTGT	60	GGGTATTGGA	70	CCATTATTGA	80
961	CCACTCCCCT	10	ATTGGTGACG	20	ATACTTTCCA	30	TACTAATCC	40	ATAACATGGC	50	TCITTTGCCAC	60	AACCTCTCTT	70	ATTGGCTATA	80
1041	TGCCAATACA	10	CTGTCTTCA	20	GAGACTGACA	30	CGGACTCTGT	40	ATTTTACAG	50	GATGGGTCT	60	CATTATTAT	70	TTACAAATTC	80
1121	ACATATACAA	10	CACCACCGTC	20	CCCAGTGCC	30	GCAGTTTTTA	40	TAAACATAAA	50	CGTGGGATCT	60	CCACGGGAAT	70	CTCGGTACG	80
1201	TGTTCCGGAC	10	ATGGGCTCTT	20	CTCCGGTAGC	30	GGCGGAGCTT	40	CTACATCCGA	50	GCCCTGTCTC	60	CATGCCCTCA	70	GCGACTCATG	80
1281	GTGCTCGGC	10	AGTCTCTTGC	20	TCCTAACAGT	30	GGAGGCCAGA	40	CTTAGGCACA	50	GCACGATGCC	60	CACCACCACC	70	AGTGTGCCGC	80
1361	ACAAGGCCGT	10	GGCGTAGGG	20	TATGTGTCTG	30	AAATGAGCT	40	CGGGGAGCGG	50	GCTTGCACCG	60	CTGACGCATT	70	TGGAAGACTT	80
1441	AAGGCAGCGG	10	CAGAAGAAGA	20	TGCAGGCAGC	30	TGAGTTGTTG	40	TGTTCTGATA	50	AGAGTCAGAG	60	GTAACCTCCG	70	TTGCGGTGCT	80
1521	GTAAACGGTG	10	GAGGGCAGTG	20	TAGTCTGAGC	30	AGTACTCGTT	40	GCTGCCGCGC	50	GCGCCACCAG	60	ACATAATAGC	70	TGACAGACTA	80
1601	ACAGACTGTT	10	CCTTTCCATG	20	GGTCTTTCT	30	GCAGTACCCG	40	TCCTTAGATC	50	TAGGTACCAG	60	ATATCAGAAT	70	TCAGTCGACA	80
1680	GCGGCCGCGA	10	TCTGTGTGC	20	CTTCTAGTTC	30	CCAGCCATCT	40	GTGTTTGCC	50	CCTCCCGCGT	60	GCCTTCTCTG	70	ACCTTGGAG	80
1761	GTGCCACTCC	10	CAGTGTCTT	20	TCCTAATAAA	30	ATGAGGAAT	40	TGCATCGCAT	50	TGCTGTAGTA	60	GGTGTCTATC	70	TATTTCTGGG	80
1841	GCTGGGGTGG	10	GGCAGCACAG	20	CAAGGGGAG	30	GATTGGGAAG	40	ACAATAGCAG	50	GCATGCTGGG	60	GATCGGTGG	70	GCTCTATGGG	80
1921	TACGGCCGCA	10	GCGGCCCTTAA	20	TTAAGGCCGC	30	AGCGGCCGTA	40	CCCAGGTGCT	50	GAAGAATTGA	60	CCCGGTCTCT	70	CGACCCGTAA	80

FIG. 1A

2/22

2001	AAAGGCGCG	TTGCTGGCGT	TTTTCCATAG	CCTCAGGAGC	ATCACAAAAA	TCGACGCICA	AGTCAGAGGT	2080
2081	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC	GCTCTCCTGT	2160
2161	CCGCTTACCG	GATACCTGTC	CGCCTTTCTC	CCCTCGGGAA	GCGTGGCGCT	TTCTCAATGC	TCACGCTGTA	2240
2241	TTCCGTTGAG	GTCGTTGCGT	CCAAGCTGGG	CTGTGTGCAC	GAACCCCGCG	TTCAGCCCGA	CCGCTGCGCC	2320
2321	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	2400
2401	AGGTATGTAG	GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA	AGGACAGTAT	2480
2481	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	CCGGCAAAAC	AACCACCGCT	2560
2561	GTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	2640
2641	CGTAATGCTC	TGCCAGTGTT	ACAACCAATT	AACCAATTCT	GATTAGAAAA	ACTCATCGAG	CATCAATGA	2720
2721	TATTCATATC	AGGATTATCA	ATACCATATT	TTTGAAAAAG	CCGTTTCTGT	AATGAAGGAG	AAAACTCACC	2800
2801	CATAGGATGG	CAAGATCCTG	GTATCGGTCT	GCGATTCCGA	CTCGTCCAAC	ATCAATACAA	CCTATTAAAT	2880
2881	AAAAATAAGG	TTATCAAGTG	AGAAATCACC	ATGAGTGACG	ACTGAATCCG	GTGAGAATGG	CAAAAGCTTA	2960
2961	TCCAGACTTG	TTCAACAGGC	CAGCCATTAC	GCTCGTCATC	AAATCACTC	GCATCAACCA	AACCGTTAAT	3040
3041	TGCGCCTGAG	CGAGACGAAA	TACGCGATCG	CTGTTAAAAG	GACAATTACA	AACAGGAATC	GAATGCAACC	3120
3121	CACITGCCAGC	GCATCAACAA	TATTTTCACC	TGAATCAGGA	TATTCITCTA	ATACCTGGAA	TGCTGTTTTC	3200
3201	CAGTGGTGAG	TAACCATGCA	TCATCAGGAG	TACGGATAAA	ATGCTTGATG	GTGGAAGAG	GCATAAATTC	3280
3281	TTTAGTCTGA	CCATCTCATC	TGTAACATCA	TTGGCAACGC	TACCTTTGCC	ATGTTTCAGA	AACAACCTCG	3360
3361	CTTCCCATAC	AATCGATAGA	TTGTGCGACC	TGATTGCCCG	ACATTATCGC	GAGCCCAATT	ATACCCATAT	3440
3441	CCATGTTTGA	ATTTAATCGC	GGCCTCGAGC	AAGACGTTTC	CCGTTGAATA	TGGCICATAA	CACCCCTTGT	3520
3521	ATGTAAGCAG	ACAGTTTTAT	TGTTTCATGAT	GATATATTTT	TATCTTGTC	AATGTAACAT	CAGAGATTTT	3600
3601	GTGGCTTTCC							3610

FIG.1B

3/22

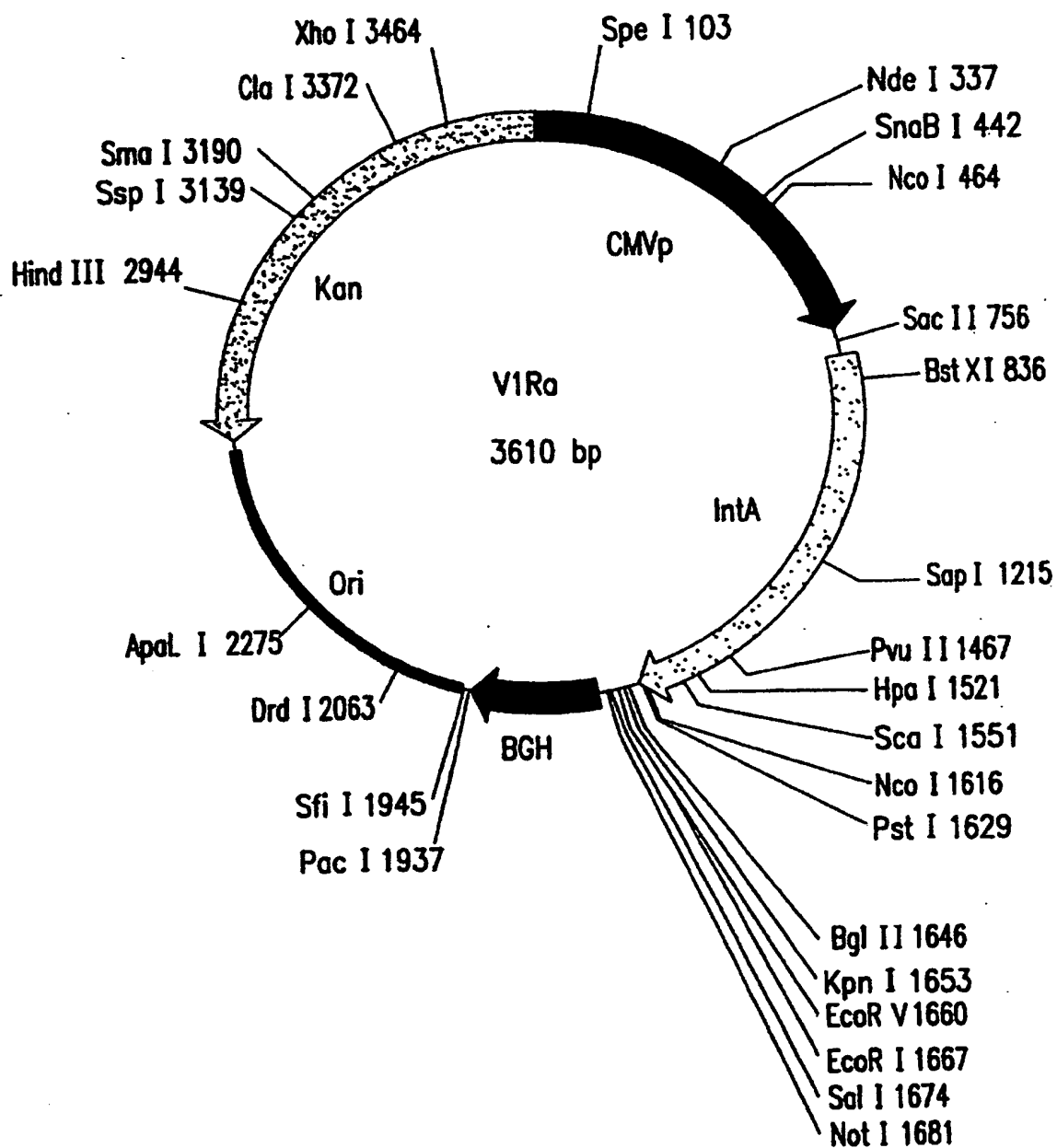


FIG.2

4/22

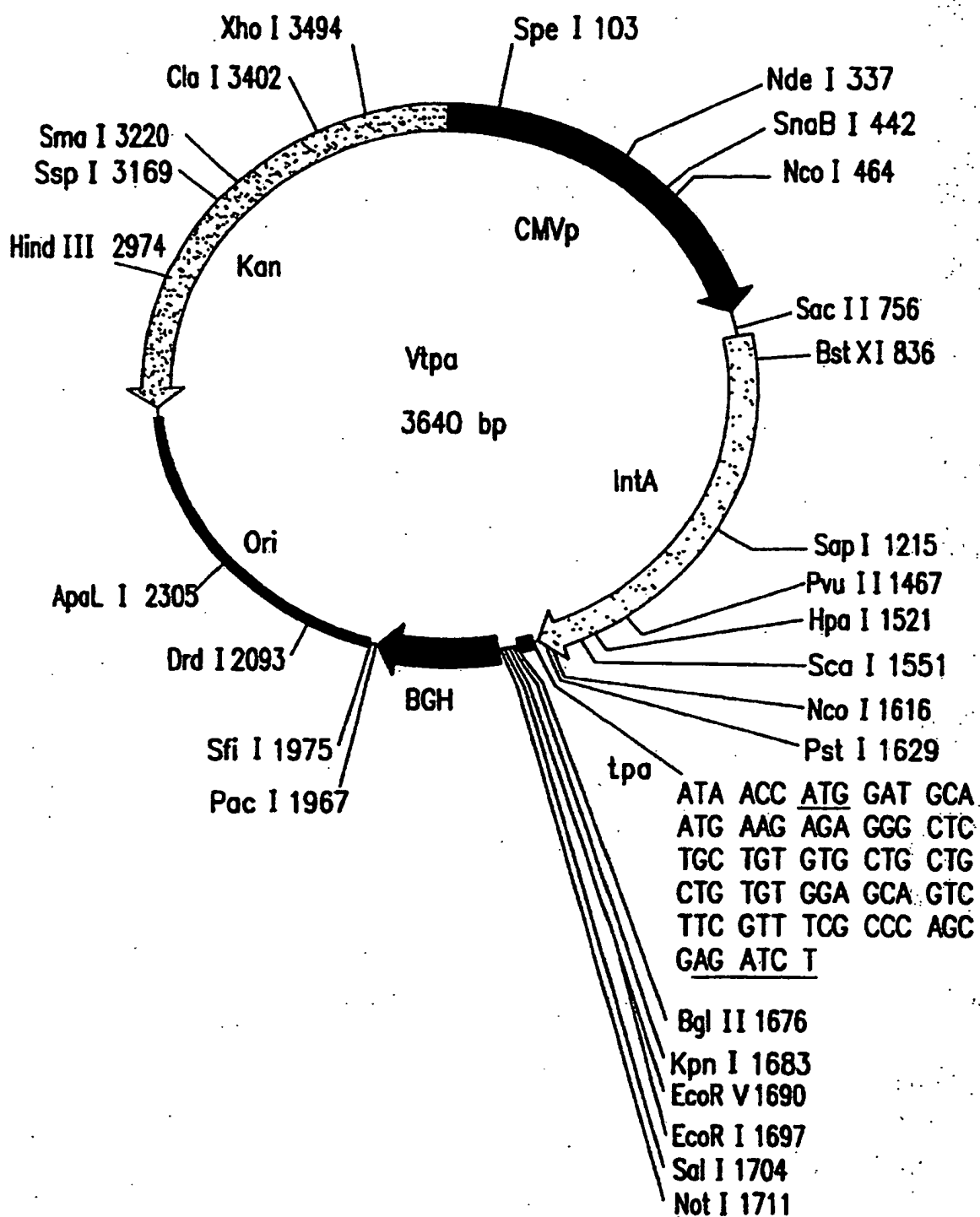


FIG.3

5/22

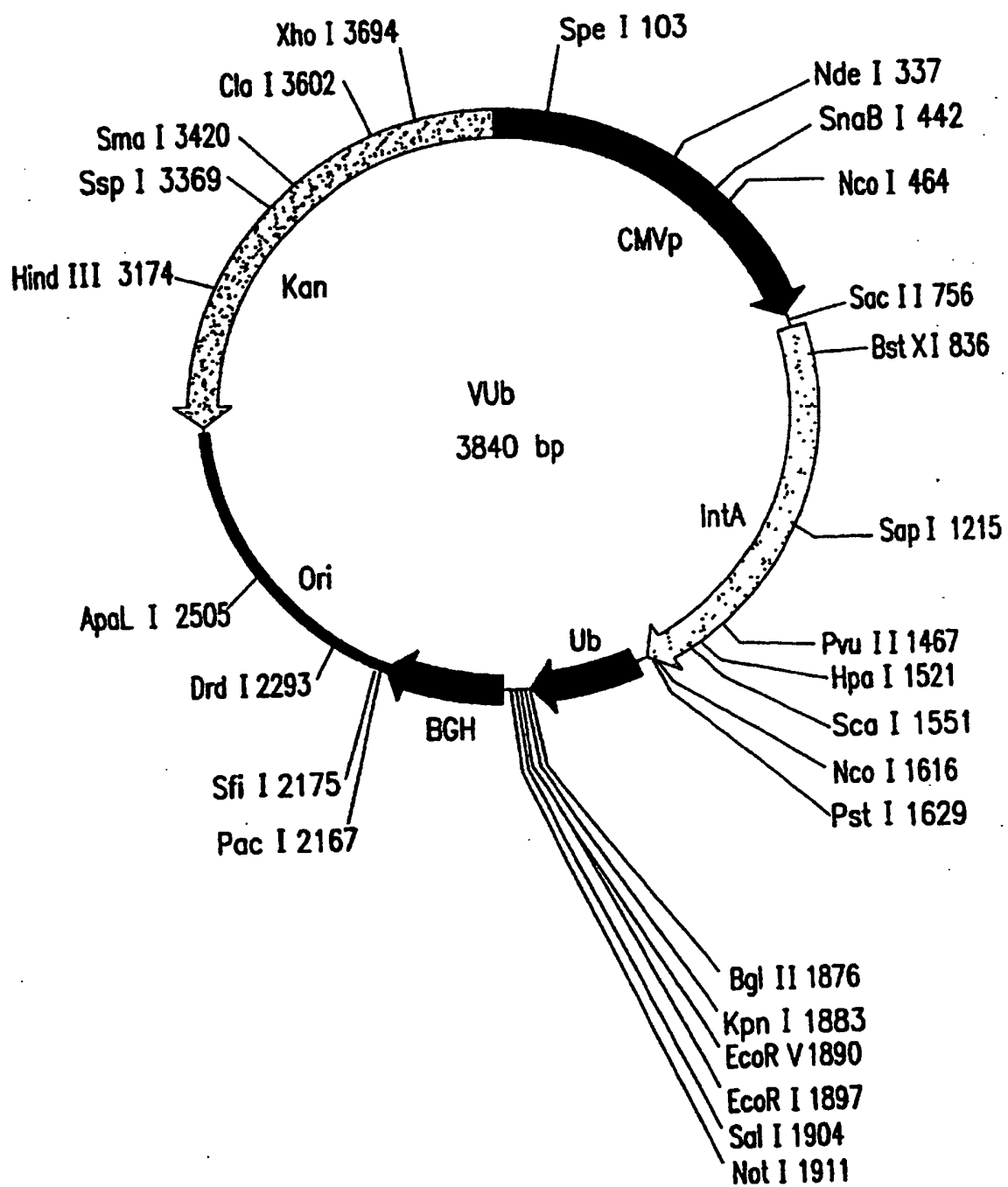


FIG.4

6/22

1/1 31/11  
 ATG AGC ACC AAC CCc AAg CCc CAg AGg AAg ACC AAg aGg AAC ACC aGg aGg CCcCAG  
 Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr asn arg arg pro gln  
 61/21 91/31  
 GAT GTg AAG TTC CCT GgG GGa GGc CAG ATT GTg GGa GGg GTc TAC cTg cTg CCc aGg AGG  
 asp val lys phe pro gly gly gln ile val gly gly val tyr leu leu pro arg arg  
 121/41 151/51  
 GGC CCC AGG cTg GgG GTg aGg Gct Acc aGg AAg ACC Tct GAG aGg Tcc Cag CCC aGg GGC  
 gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly  
 181/61 211/71  
 AGG aGg CAG CCc ATC CCC AAG GcC aGg aGg CCT GAG GGC cGc TCC TGG GCC CAG CCT GGC  
 arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly  
 241/81 271/81  
 TAC CCC TGG CCC CTg TAT GGC AAT GAa GGC TGG Gct GGC TGG CTg CTg TCC CCC  
 try pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro  
 301/101 331/111  
 aGg GGC Tcc aGg CCc tcc TGG GGC CCC ACa GAC CCC aGg aGg aGg Tcc aGg AAC cTg GGC  
 arg gly ser arg pro ser trp gly pro thr asp pro arg arg arg ser arg asn leu gly  
 361/121 391/131  
 AAg GTg ATT GAC ACC CTg ACc Tgt GGC Ttt Gct GAC CTg ATG GGC TAC ATC CCc CTg GTg  
 lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val  
 421/141 451/151  
 GgG Gct CCT GTg GGa GgG GTg Gct AGG Gct CTG Gct CAT GgG GTg AGG GTg CTG GAG GAT  
 gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp  
 481/161 511/171  
 GGG GTG AAC TAT Gct ACT GGC AAC cTg CCT GGC TGC Tcc TTC Tcc ATC TTC CTg CTG GCC  
 gly val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala  
 541/181 571/191  
 CTG CTc TCC TGC CTG ACa GTg CCT GCT TCT GCC  
 leu leu ser cys leu thr val pro ala ser ala

FIG. 5

7/22

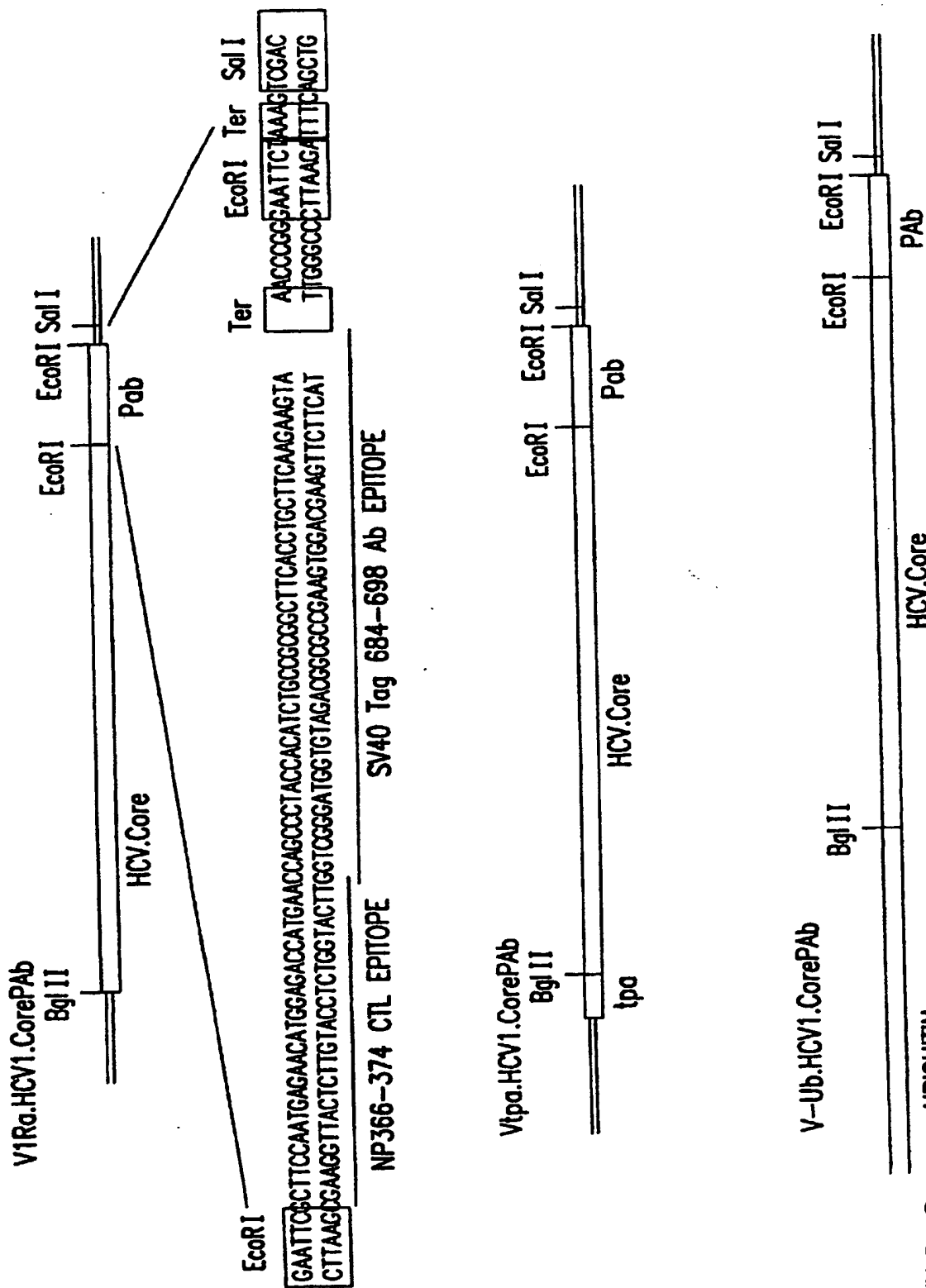


FIG. 6

8/22

1/1 31/11  
 ATG AGC AGC AAT CCT AAA CCA AGA AAA ACC AAA CGT AAC ACC AAC CGC CGC CCA CAG  
 Met ser thr asn pro lys pro gln arg lys thr lys arg asn thr arg arg pro gln  
 61/21 91/31  
 GAC GTC AAG TTC CCg GGC GGT CAG ATC GTT GGT GGA GTT TAC TTC TTT CCG CGC AGG  
 asp val lys phe pro gly gly gln ile val gly val tyr leu leu pro arg arg  
 121/41 151/51  
 GGC CCC AGG TTG GGT GTG CGC GCG ACT aGg AAG ACT Tcc GAG CGG TCG CAA CCT CGT GGA  
 gly pro arg leu gly val arg ala thr arg lys thr ser glu arg ser gln pro arg gly  
 181/61 211/71  
 AGG CGa CAG CCT ATC CCC AAG Gct CGc CGG CCC GAG Ggc AGG TCC TGG GCT CAG CCC GGG  
 arg arg gln pro ile pro lys ala arg arg pro glu gly arg ser trp ala gln pro gly  
 241/81 271/91  
 TAC Cct TGG CCc CTC TAt GGC AAT GAg GGC Ttc GGG TGG GCA GGA TGG CTC CTG TCC CCC  
 tyr pro trp pro leu tyr gly asn glu gly phe gly trp ala gly trp leu leu ser pro  
 301/101 331/111  
 CGC GGC TCT CGg Cct agT TGG GGC CCC Act GAC CCC CGG CGt AGG TCG CGC AAT TTG GGT  
 arg gly ser arg pro ser trp gly pro thr asp pro arg arg ser arg asn leu gly  
 361/121 391/131  
 AAG GTC ATC GAT ACC CTC ACG TGC GGC TTC GCC GAC CTC ATG GgG TAC ATC CCG CTC GTC  
 lys val ile asp thr leu thr cys gly phe ala asp leu met gly tyr ile pro leu val  
 421/141 451/151  
 GGC GCC CCc GTa GgG GGC GTC GCC Agg GCC CTG GCG CAT GGC GTC AGG GtT cTG GAG GAC  
 gly ala pro val gly gly val ala arg ala leu ala his gly val arg val leu glu asp  
 481/161 511/171  
 GGG gtg AAC TAT GCA ACA GGG Aat tTg cCc GGT TGC TCT TTC TCT ATC TTC CTC cTG Gct  
 glu val asn tyr ala thr gly asn leu pro gly cys ser phe ser ile phe leu leu ala  
 541/181 571/191  
 CTg CTg Tcc TGC CTG ACC GTC CCA Gct TCT GCT  
leu leu ser cys leu thr val pro ala ser ala

FIG. 7

9/22

TABLE 3  
CODON UTILIZATION IN HUMAN PROTEIN-CODING SEQUENCES

a	b	c	d	e	f	a	b	c	d	e	f
F	UUU UUC	68 125	0.35 0.65	193	4.5	Y	UAU UAC	72 81	0.47 0.53	153	3.6
L	UUA UUG CUU CUC CUA CUG	20 42 50 99 30 204	0.05 0.09 0.11 0.22 0.07 0.46	445	10.4	H	CAU CAC	44 61	0.42 0.58	105	2.5
I	AUU AUC AUA	28 79 16	0.23 0.64 0.13	123	2.9	Q	CAA CAG	50 142	0.26 0.74	192	4.5
M	AUG	77	1.00	77	1.8	N	AAU AAC	51 97	0.34 0.66	148	3.5
V	GUU GUC GUA GUG	35 72 25 134	0.13 0.27 0.09 0.50	266	6.2	K	AAA AAG	137 166	0.45 0.55	303	7.0
S	UCU UCC UCA UCG AGU AGC	59 91 37 25 37 100	0.17 0.26 0.11 0.07 0.11 0.29	349	8.1	D	GAU GAC	79 130	0.38 0.62	209	4.9
P	CCU CCC CCA CCG	51 86 51 24	0.24 0.41 0.24 0.11	212	4.9	E	GAA GAG	125 186	0.40 0.60	311	7.3
T	ACU ACC ACA ACG	47 113 50 28	0.20 0.47 0.21 0.12	238	5.6	C	UGU UGC	44 103	0.30 0.70	147	3.4
A	GCU GCC GCA GCG	91 119 51 37	0.31 0.40 0.17 0.12	298	7.0	W	UGG	56	1.00	56	1.3
						R	CGU CGC CGA CGG AGA AGG	19 40 22 33 51 50	0.09 0.19 0.10 0.15 0.24 0.23	215	5.0
						G	GGU GGC GGA GGG	36 108 42 59	0.15 0.44 0.17 0.24	245	5.7

TOTAL 4285 RESIDUES EXCLUDING  
N-TERMINAL METHIONINE RESIDUES

FIG.8

10/22

1/1  
 atg TAT GAG GTG aGg AAT GTc Tct Ggc GTc TAC CAT GTg ACc AAT GAC TGC TCC AAC Tcc  
 M Y E V R N V S G V Y H V T N D C S N S  
 31/11  
 61/21  
 tGc ATT GTc TAT GAG Gct Gct GAC ATG ATc ATG CAC ACC Cct Ggc Tgt GTg CCa Tgt GTg  
 C I V Y E A A D M I M H T P G C V P C V  
 91/31  
 121/41  
 aGg GAG Ggc AAC TCC TCC aGg TGC TGG GTg Gcc CTg ACC CCC ACc CTg Gct GCC AGG AAC  
 R E G N S S R C W V A L T P T L A A R N  
 151/71  
 181/61  
 tcC tcC ATC CCC ACC ACC ATc aGg aGg CAT GTg GAC cTg CTg GTg Ggc Gct GCT GCC  
 S S I P T T I R R H V D L L V G A A A  
 211/71  
 241/81  
 CTg TGC Tct GCC ATG TAT GTG Ggc GAC CTg TGT Ggc TCT GTc TTC CTg GTg TCC CAG gTg  
 L C S A M Y V G D L C G S V F L V S Q L  
 271/91  
 301/101  
 TTC ACC TTC TCC CCc aGg aGg TAT GAG Act GTg CAG GAC TGC AAC TGC Tcc CTg TAC CCT  
 F T F S P R R Y E T V Q D C N C S L Y P  
 331/111  
 361/121  
 GGC CAT GTc Tct Ggc CAC aGg ATG Gcc TGG GAC ATG ATG ATG AAC TGG Tcc CCc ACc Act  
 G H V S G H R M A W D M M N W S P T T  
 391/131  
 421/141  
 GCC cTg GTG GTc Tcc CAG cTg CTg aGg ATt CCc CAG Gct GTg GTG GAC ATG GTG TGT GgG  
 A L V V S Q L L R I P Q A V V D M V V G  
 451/151  
 481/161  
 GCC CAC TGG Ggc GTg CTg Gct Ggc CTg GCC TAC TAC TCC ATG GTG Ggc AAC TGG GCC AAG  
 A H W G V L A G L A Y Y S M V G N W A K  
 511/171  
 541/181  
 GTg cTg ATT GTG ATG CTg CTg TTT Gct Ggc GTg GAT Ggc taa  
 V L I V M L L F A G V D G \*

FIG. 9

11/22

1/1  
 atg ACC ACC TAT GTc Tct GTG Ggc CAT Gcc tcC CAG ACC ACC aGg aGg GTg Gcc TCC TTC  
 M T T Y V S V G H A S Q T T R R V A S F  
 31/11  
 61/21  
 TTC tcc CCT GGC Tct Gcc CAG AAg ATC CAg CTg GTg AAC ACC AAt GGC tcc TGG CAC ATC  
 F S P G S A Q K I Q L V N T N G S W H I  
 91/31  
 121/41  
 AAC AGG ACT GCC CTG AAt TGC AAt GAG TCC ATC AAC ATC Ggc TTC TTT Gct Gcc CTG TTC  
 N R T A L N C N E S I N T G F F A A L F  
 151/51  
 181/61  
 TAT GTg AAG AAG TTC AAC TCC TCT Ggc TGC Tct GAG aGg ATG Gcc tct Tgc aGg CCC ATT  
 Y V K K F N S S G C S E R M A S C R P I  
 211/71  
 241/81  
 GAC AGG TTt Gcc CAG Ggc TGG Ggc CCC ATC ACC CAT GCT GAG Tcc aGg tcc Tct GAC CAg  
 D R F A Q G G W G P I T H A E S R S S D Q  
 271/91  
 301/101  
 TAC TGC TGG CAC TAT GCC CCC CAg CCa TGT Ggc ATt GTG CCT Gcc cTG CAT GTc  
 R P Y C W H Y A P Q P C G I V P A L H V  
 331/111  
 361/121  
 Tgt Ggc CCT GTc TAc Tgc TTC ACC CCa tcC CCT GTg GTg GTG Ggc ACg ACT GAC aGg TTt  
 C G P V Y C F T P S P V V V G T T D R F  
 391/131  
 421/141  
 GGC GTg CCC ACC TAC AAC TGG Ggc GAC AAt GAG ACT GAT GTG CTg CTg AAC AAC ACC  
 G V P T Y N W G D N E T D V L L L N N T  
 451/151  
 481/161  
 aGg CCC CCC CAg Ggc AAC TGG TTt Ggc Tgc ACC TGG ATG AAC tcC ACT Ggc TTC ACC AAG  
 R P P Q G N W F G C T W M N S T G F T K  
 511/171

FIG.10A

12/22

541/181 571/191  
 ACC Tgt Ggc Ggc CCC CCa Tgc Aac Att Ggc Ggc Gct Ggc AAC ACC cTG ACC Tgc CCC  
 T C G G P P C N I G G A G N N T L T C P  
 601/201 631/211  
 Act Gac Tgc TTC aGg Aag Cat Cct GAG GCC Acc TAC ACC AAG Tgt Ggc Tct Ggc CCa Tgg  
 T D C F R K H P E A T Y T K C G S G P W  
 661/221 691/231  
 cTG Acc CCc AGG Tgc ATg GTg Gac TAC CCa TAC Agg CTg Tgg CAC TAC CCa Tgc Acc TTC  
 L T P R C M V D Y P Y R L W H Y P C T F  
 721/241 751/251  
 AAC TTC ACC ATC TTc AAG ATC AGG ATg TAT GTg Ggc Ggc GTg GAG CAC AGG CTg Aat Gct  
 N F T I F K I R M Y V G G V E H R L N A  
 781/261 811/271  
 Gcc Tgc AAC Tgg Acc aGg Ggc GAG aGg Tgc AAC Att GAG GAC AGG GAC AGG Tct GAG CTg  
 A C N W T R G E R C N I E D R D R S E L  
 841/281 871/291  
 tcc CCc CTg CTg CTg Tcc Acc Act GAG Tgg Cag ATc CTg CCa Tgc Tcc TTc ACC ACC CTg  
 S P L L L S T T E W Q I L P C S F T T L  
 901/301 931/311  
 CCT Gcc CTg Tcc ACT Ggc cTG ATC Cat CTg Cat Cag AAC Att GTg Gat GTg Cag TAC CTg  
 P A L S T G L I H L H Q N I V D V Q Y L  
 961/321 991/331  
 TAT Ggc GTg Ggc Tct Gct GTg GTc Tcc ATT GTg ATC Aag Tgg GAG TAT GTg CTg CTg CTg  
 Y G V G S A V V S I V I K W E Y V L L L  
 1021/341  
 TTC CTg CTg CTg Gct Gat Gcc taa  
 F L L L A D A \*

FIG. 10B

13/22

```

1/1      31/11
atg TAT GAG GTG aGg AAt GTc Tct Ggc GTc TAc CAT GTg Acc AAt GAc Tgc Tcc AAC Tcc
M Y E V R N V S G V Y H V T N D C S N S

61/21    91/31
tGc ATT GTc TAT GAG GcT GcT GAc ATg ATc ATg CAC ACC Cct Ggc Tgt GTg CCa Tgt GTg
C I V Y E A A D M I M H T P G C V P C V

121/41   151/51
aGg GAG Ggc AAC Tcc Tcc aGg Tgc Tgg GTg Gcc CTg Acc CCC Acc CTg Gct GCC AGG AAC
R E G N S S R C W V A L T P T L A A R N

181/61   211/71
tcc tcc ATc CCC Acc Acc ATc aGg aGg CAT GTg GAc cTg CTg GTg Ggc Gct Gct Gcc
S S I P T T I R R H V D L L V G A A A

241/81   271/91
CTg Tgc Tct Gcc ATg TAt GTg Ggc GAc CTg Tgt Ggc Tct Gtc TTC CTg GTg Tcc Cag cTg
L C S A M Y V G D L C G S V F L V S Q L

301/101  331/111
TTC Acc TTC Tcc Ccc aGg aGg TAT GAG Act GTg CAG GAc Tgc AAC Tgc Tcc CTg TAc Cct
F T F S P R R Y E T V Q D C N C S L Y P

361/121  391/131
GGC CAT GTc Tct Ggc CAC aGg ATg Gcc Tgg GAc ATg ATg ATg AAC Tgg Tcc Ccc Acc Act
G H V S G H R M A W D M M N W S P T T

421/141  451/151
GCC cTg GTg GTc Tcc CAG cTg CTg aGg ATt Ccc Cag Gct GTg GTg GAc ATg GTg GTg Ggc
A L V V S Q L L R I P Q A V V D M V V G

481/161  511/171
GCC CAC Tgg Ggc GTg CTg Gct Ggc CTg GCC TAc TAc Tcc ATg GTg Ggc AAC Tgg Gcc AAG
A H W G V L A G L A Y Y S M V G N W A K

```

FIG.11A

14/22

541/181 571/191  
 GTg cTg ATT GTG ATG CTg CTg TTT Gct GGC GTg GAT Ggc Acc ACC TAT GTc Tct GTG Ggc  
 V L I V M L L F A G V D G T T Y V S V G  
 601/201 631/211  
 CAT Ggc tcC CAG ACC ACC aGg aGg GTg Ggc TCC TTC TTC tcc CCT GGC Tct Gcc CAG AAG  
 H A S Q T T R R V A S F F S P G S A Q K  
 661/221 691/231  
 ATC Cag CTg GTg AAC ACC AAT GGC tcc TGG CAC ATC AAC AGG ACT GCC CTG Aat TGC Aat  
 I Q L V N T N G S W H I N R T A L N C N  
 721/241 751/251  
 GAG TCC ATC AAC ACT GGC TTC TTT Gct Gcc CTG TTC TAT GTg AAG AAG TTC AAC Tcc TCT  
 E S I N T G F F A A L F Y V K K F N S S  
 781/261 811/271  
 Ggc TGC Tct GAG aGg ATG Gcc tct Tgc agg CCC ATT GAC AGG Ttt Gcc Cag Ggc TGG Ggc  
 G C S E R M A S C R P I D R F A Q G W G  
 841/281 871/291  
 CCC ATC ACC CAT GCT GAG Tcc aGg tcC Tct GAC Cag AGG CCa TAC TGC TGG CAC TAT Gcc  
 P I T H A E S R S S D Q R P Y C W H Y A  
 901/301 931/311  
 CCC Cag CCa TGT Ggc ATt GTG CCT Gcc cTg Cat GTc Tgt Ggc CCT GTc TAc Tgc TTC ACC  
 P Q P C G I V P A L H V C G P V Y C F T  
 961/321 991/331  
 CCa tcC CCT GTg GTg Ggc Acc Act GAC agg Ttt GGC GTg CCC Acc TAc AAC TGG Ggc  
 P S P V V V G T T D R F G V P T Y N W G  
 1021/341 1051/351  
 GAC AAT GAG Act GAT GTG CTg CTg CTg AAC AAC Acc aGg CCC Ccc Cag Ggc AAC TGG TTe  
 D N E T D V L L L L N N T R P P Q G N W F

FIG. 11B

15/22

1081/361  
 GGC Tgc Acc TGG ATG AAC tcc Act Ggc TTC ACC AAG Acc Tgt Ggc GGC CCC Cca Tgc AAC  
 G C T W M N S T G F T K T C G G P P C N  
 1111/37  
 1141/381  
 ATT GGC Ggc Gct GGC AAC ACC cTG ACC TGC CCC ACT GAC TGC TTC aGG AAG CAT Cct  
 I G G A G N N T L T C P T D C F R K H P  
 1171/391  
 1201/401  
 GAG GCC Acc TAC ACC AAG Tgt Ggc Tct Ggc Cca TGG cTG ACC CCC AGG TGC ATG GTg GAC  
 E A T Y T K C G S G P W L T P R C M V D  
 1231/411  
 1261/421  
 TAC Cca TAC AGg CTg TGG CAC TAC Cca TGC Acc TTC AAC TTC ACC ATC TTC AAG ATC AGG  
 Y P Y R L W H Y P C T F N F T I F K I R  
 1291/431  
 1321/441  
 ATG TAT GTG Ggc Ggc GTG GAG CAC AGG CTg Aat Gct Gcc TGC AAC TGG ACC aGg Ggc GAG  
 M Y V G G V E H R L N A A C N W T R G E  
 1351/451  
 1381/461  
 aGg Tgc AAC ATg GAG CAC AGG GAC AGG Tct GAG CTg tcc CCC CTG CTg CTG Tcc Acc ACT  
 R C N I E D R D R S E L S P L L L S T T  
 1411/471  
 1441/481  
 GAG TGG CAG ATc CTg Cca Tgc TCC TTC ACC ACTg Cct Ggc CTG TCC ACT Ggc cTG ATC  
 E W Q I L P C S F T T L P A L S T G L I  
 1471/491  
 1501/501  
 CAT CTg CAT CAG AAC ATt GTG GAT GTG Cag TAC CTG TAT Ggc GTg Ggc Tct Gct GTg GTc  
 H L H Q N I V D V Q Y L Y G V G S A V V  
 1531/511  
 1561/521  
 TCC ATT GTG ATC AAg TGG GAG TAT GTg CTG CTG CTg TTC CTg CTg CTg Gct Gat Gcc taa  
 S I V I K W E Y V L L L F L L L A D A \*

FIG.11C

1/1  
 atg Tct Ggc Tcc Tgg CTg AGg GAT GTc TGG GAC TGG ATc TGC ACT GTG cTG ACT GAC TCC  
 M S G S W L R D V W D W I C T V L T D F  
 31/11  
 61/21  
 AAG ACC TGG CTg CAT TCC AAG CTg CTG CCc aGg CTG CCt Ggc GAC CCa TTC TTC Tcc TGc  
 K T W L H S K L L P R L P G D P F F S C  
 91/31  
 121/41  
 CAg aGg Ggc TAC AGg Ggc GTc TGG aGg Ggc GAT Ggc GTg ATG CAg ACC ACC TGC CCa TGT  
 Q R G Y R G V W R G D G V M Q T T C P C  
 151/51  
 181/61  
 Ggc GCC CAG ATC ACT Ggc CAT GTg AAg AAT Ggc TCC ATG AGg ATt GTg Ggc CCc AAg ACC  
 G A Q I T G H V K N G S M R I V G P K T  
 211/71  
 241/81  
 TGc tCc AAC ACC TGG CAT Ggc ACC TTC CCC ATC AAT GCC TAC ACC ACT Ggc CCa TGC ACC  
 C S N T W H G T F P I N A Y T T G P C T  
 271/91  
 301/101  
 CCa TCC Cct Gcc CCC AAC TAC TCC AGg Gcc CTG TGG aG GTG GCT GCT GAG GAG TAT GTG  
 P S P A P N Y S R A L W R V A A E Y V  
 331/111  
 361/121  
 GAg GTg ACC aGg GTG Ggc GAC TTC CAC TAT GTG ACT Ggc ATG ACC ACT GAC AAt GTg AAg  
 E V T R V G D F H Y V T G M T T D N V K  
 391/131  
 421/141  
 TGC CCa TGC CAG GTg Cct GCC Cct GAg TTC TTC ACT GAg GTG GAT Ggc GTG aGg cTG CAC  
 C P C Q V P A P E F T E V D G V R L H  
 451/151  
 481/161  
 AGG TAT GCC Cct GCC TGC AAg CCc CTg CTg aGg GAT GAg GTg ACC TTC CAG GTg Ggc CTg  
 R Y A P A C K P L L R D E V T F Q V G L  
 511/171

FIG. 12A

17/22

541/181 AAC Cag TTC Cct GTg Ggc Tcc Cag CTg CCa TGT GAG CCT GAg CCT GAT GTg ACT GTG CTg  
 N Q F P V G S Q L P C E P E P D V T V L  
 571/191  
 601/201 ACC TCC ATG CTg ACT GAg CCa TCC CAC ATc ACT Gct GAG ACT Gcc AAG aGg AGG CTG GCC  
 T S M L T E P S H I T A E T A K R L A  
 631/211  
 661/221 AGg Ggc Tcc Cct CCa TCC CTg GCC tcc Tcc TGcc tcc CAG cTG TCT Gct CCa Tcc cTG  
 R G S P P S L A S S S A S Q L S A P S L  
 691/231  
 721/241 AAG Gcc ACC TGC ACC ACC aGg CAT GAC TCC Cct GAT Gct GAC CTg ATt GAG GCC AAC CTg  
 K A T C T T R H D S P D A D L I E A N L  
 751/251  
 781/261 CTG TGG aGg CAG GAG ATG GGC Ggc AAC ATc ACC aGg GTG GAG Tct GAG AAC AAG GTg GTg  
 L W R Q E M G G N I T R V E S E N K V V  
 811/271  
 841/281 ATc CTg GAC Tcc TTT GAg CCC CTg aGg Gct GAG GAG GAT GAG AGG GAg GTc Tct GTG Gct  
 I L D S F E P L R A E E D E R E V S V A  
 901/301 Gct GAG ATc CTg aGg AAg tcc AGG AAG TTC CCC Cct Gcc cTG CCC ATc TGG GcG aGg CCa  
 A E I L R K S R K F P P A L P I W A R P  
 931/311  
 961/321 tcc TAC AAC CCa CCc CTg CTg GAG Tcc TGG AAG GAC Cct GAC TAT GTg CCC Cct GTG GTg  
 S Y N P P L L E S W K D P D Y V P P V V  
 991/331  
 1021/381 Cat Ggc TGC CCC CTG CCc ACC ATG GCC CCa CCc ATc CCc CCc aGg AGG AAG AGG  
 H G C P L P P T M A P P I P P R R K R  
 1051/371

FIG. 12B

18/22

1081/361  
 ACT GTg GTg CTG ACT GAg TCC ACT GTc TCC TCT GCC cTG GCT GAG CTg GCC ACC AAG ACC  
 T V V L T E S T V S S A L A E L A T K T  
 1111/371  
 1141/381  
 TTC GCC tcC Tct TCC Tct Gct GTg GAC tct GGC Act GCc ACG GCC CCc CCT GAC CAG  
 F G S S G S A V D S G T A T A P P D Q  
 1171/391  
 1201/401  
 CCa Tct GAT GAT GGC GAC AGg GGC Tct GAT GAT GAG TCC TAC TCC TCC ATG CCC CCC CTg  
 P S D D G D R G S D D E S Y S S M P P L  
 1231/411  
 1261/421  
 GAG GGC GAG CCT GGC GAC CCT GAC CTg tct GAT GGC TCC TGG TCC Act GTc tct GAG GAG  
 E G E P G D P D L S D G S W S T V S E E  
 1291/431  
 1321/441  
 GCc tct GAG GAT GTg GCc TGC TGC TCC taa  
 A S E D V A C C S \*

FIG.12C

19/22

1/1 31/11  
 ATG Tcc TAC Acc TGG Act GGC GCC CTg ATC Acc CcA Tgt Gct Gct GAG GAG tcc AAG CTG  
 M S Y T W T G A L I T P C A A E E S K L  
 61/21 91/31  
 CCC ATC AAC Ccc cTG tcc AAC Tcc cTG CTG agg CAT CAC AAC ATG GTc TAt GCC Acc Acc  
 P I N P L S N S L L R H H N M V Y A T T  
 121/41 151/51  
 TCC agg tct Gct GGC CTg agG CAG Aag AAG GTg ACC TTT GAC AGg CTG CAT GTg Cct Gat  
 S R S A G L R Q K K V T F D R L H V P C  
 181/61 211/71  
 GAC CAC TAC agG Gat GTG CTg AAG GAG ATG AAG GCC AAG Gcc TCC Act GTg AAG Gcg Aag  
 D H Y R D V L K E M K A K A S T V K A K  
 241/81 271/91  
 CTg CTg TCT GTg GAg GAg GCC TGC AAG CTG Acc Cct Ccc CAC Tct GCC AGg TCC Aag TTT  
 L L S V E E A C K L T P P H S A R S K F  
 301/101 331/111  
 GGC TAT GGC GCC AAG Gat GTg agG AAC CTg TCC tcc AAG Gct GTg AAC CAC ATC CAC Tct  
 G Y G A K D V R N L S S K A V N H I H S  
 361/121 391/131  
 GTc TGG AAG GAC cTG CTG GAg GAC ACT GAg Acc Ccc ATT GAC Acc ACC ATC ATG GCC Aag  
 V W K D L L E D T E T P I D T T I M A K  
 421/141 451/151  
 AAT GAG GTc TTC TGT GTg CAg Cct GAG Aag GGC GGC agG AAG Cct GCC agg CTg ATt GTc  
 N E V F C V Q P E K G G R K P A R L I V  
 481/161 511/171  
 TTC Cct GAg CTg Ggc GTg agg GTG Tgt GAG Aag ATG GCC CTg TAt Gat GTG GTc TCC Acc  
 F P E L G V R V C E K M A L Y D V V S T

FIG.13A

20/22

541/181 CAG Gct GTG ATG GGC TCC TCC TAT GGC TTC CAG TAC TCC CCT GGC CAG aGG GTg  
 L P Q A V M G S S Y G F Q Y S P G Q R V  
 571/191  
 601/201 GAG TTC CTG GTG AAT GCC TGG Aag TCC AAG Aag AAC CCC ATG GGC TTt GCC TAC TGC ACC  
 E F L V N A W K S K K N P M G F A Y C T  
 631/211  
 661/221 TTT GAC TCC Act GTg ACT GAG tct GAC ATC aGg GTg GAG GAG TCC ATc TAC CAG  
 aGg Tgc R C F D S T V T E S D I R V E E S I Y Q  
 691/231  
 721/241 GAC cTG Gct Cct GAG GCC Agg CAG GTg ATc AGG TCC CTg ACT GAG aGG CTg TAC  
 C C D L A P E A R Q V I R S L T E R L Y  
 751/251  
 781/261 GGC CCC CTG Acc AAC TCC AAG Ggc CAG AAC Tgt Ggc TAC aGg aGG TGC aGg GCC  
 I G G P L T N S K G Q N C G Y R C R A  
 811/271  
 841/281 GTG CTG Acc ACT AAC Tgt Ggc AAC ACC CTg Acc Tgc TAC cTG AAG GCC TCT Gct  
 tct Ggc S G V L T T N C G N T L T C Y L K A S A  
 871/291  
 901/301 GCT GcC aGg GCT GCC AAG CTg Cat GAC TGC ACC ATG CTg GTc Tgt Ggc Gat GAC CTg GTg  
 A C R A A K L H D C T M L V C G D D L V  
 931/311  
 961/321 TGT GAg tct Gct Ggc ACC CAG GAG Gat Gct Gcc tcC CTg aGg GTc TTC Act GAG  
 V I C E S A G T Q E D A A S L R V F T E  
 991/331  
 1021/341 AGG TAC TCT GCC CCc Cct Ggc GAC Cct CCC Cag Cct GAg TAT GAC cTG GAG  
 A M T R Y S A P P G D P P Q P E Y D L E  
 1051/351

FIG. 13B

21/22

1081/361  
 cTg ATc Acc Tcc Tgc Tcc Tcc Aat Gtc Tct Gtg Gcc Cat Gat Gcc Tct Ggc Aag agg Gtc  
 L I T S C S S N V S V A H D A S G K R V  
 1111/371  
 1141/381  
 TAC TAC CTg Acc agg GAC CCC ACC ACC CCC CTg Gcc AGG Gct Gcc TGG GAg Act Gcc AGg  
 Y Y L T R D P T T P L A R A A W E T A R  
 1171/391  
 1201/401  
 CAC ACC Cct GTg AAC TCC TGG CTg GGC AAC ATC ATC ATG TAT Gcc CCC ACC CTG TGG Gcc  
 H T P V N S W L G N I I M Y A P T L W A  
 1231/411  
 1261/421  
 AGG ATG ATc CTg ATG ACC CAC TTC TTC TCC ATC CTg CTg Gcc CAG GAG CAG CTg GAg AAg  
 R M I L M T H F S I L L A Q E Q L E K  
 1291/431  
 1321/441  
 GCC CTG Ggc Tgc CAG ATt TAT Ggc GCC Acc TAC TTC ATT GAg CCC CTg GAC CTg CCC CAG  
 A L G C Q I Y G A T Y F I E P L D L P Q  
 1351/451  
 1381/461  
 ATC ATc CAG agg CTg CAT Ggc CTg tct Gcc Ttc Tcc CTg CAC tcc TAC Tcc Cct Ggc GAg  
 I I Q R L H G L S A F S L H S Y S P G E  
 1411/471  
 1441/481  
 ATC AAC AGG GTG Gcc Tcc Tgc CTg AGG AAg CTg Ggc GTg Ccc CCC cTg aGg GTg TGG AGg  
 I N R V A S C L R K L G V P P L R V W R  
 1471/491  
 1501/501  
 GAc aGG GCC AGg tct GTg aGg Gcc AAG CTg CTG TCC CAG Ggc Ggc AGG Gct GCC Acc TGT  
 H R A R S V R A K L L S Q G G R A A T C  
 1531/511  
 1561/521  
 GGC AAG TAC CTg TTC AAC TGG Gct GTG AGG ACC AAG CTg AAg CTg Acc Ccc ATc Cct GCT  
 G K Y L F N W A V R T K L K L T P I P A  
 1591/531

FIG.13C

22 / 22

1621/541  
 GCC TCC CAG cTg GAC cTg Tct GGC TGG Ttt GTg GCT Ggc TAC tct Ggc Ggc GAC ATc TAc  
 A S Q L D L S G W F V A G Y S G G D I Y  
 1681/561  
 CAC tcc CTG Tcc aGg GCC aGg CCC aGg TGG TTC ATG TGG TGC CTg CTg CTg TCT GTg  
 H S L S R A R P R W F M W C L L L L S V  
 1741  
 GGc GTg GGC ATC TAC CTG CTg CCC AAC aGg TGA  
 G V G I Y L L P N R \*

FIG. 13D

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : A61N 43/04; C12Q 1/68; C12N 15/00; C07H 21/02; A61K 39/00  
US CL : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1  
According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 514/44; 435/6, 320.1; 536/23.1; 434/184.1, 192.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
none

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
APS, STN, MEDLINE, BIOSIS, EMBASE, CAPLUS, WPIDS, SCISEARCH

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Selby et al. Expression, identification and subcellular localization of the proteins encoded by the hepatitis C viral genome. Journal of General Virology. 1993. Vol. 74, pages 1103-1113, see entire document.	1-3
X	Bukh et al. Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proc. Natl. Acad. Sci. August 1994. Vol. 91, pages 8239-8243, see entire document.	1-3
Y	Lathe. Synthetic Oligonucleotide Probes Deduced from Amino Acid Sequence Data Theoretical and Practical Considerations. J. Mol. Biol. 1985. Vol. 183, pages 1-12, see entire document.	1-3

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

\*A\* document defining the general state of the art which is not considered to be of particular relevance

\*B\* earlier document published on or after the international filing date

\*L\* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

\*O\* document referring to an oral disclosure, use, exhibition or other means

\*P\* document published prior to the international filing date but later than the priority date claimed

\*T\*

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\*

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\*

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

\*A\*

document member of the same patent family

Date of the actual completion of the international search

28 AUGUST 1997

Date of mailing of the international search report

11 SEP 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

ANDREW WANG

Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	Grantham et al. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Research. 1981. Vol. 9, No. 1, pages r43-r74, see entire document.	1-3
A, P	Ide et al. Characterization of the nuclear localization signal and subcellular distribution of hepatitis C virus nonstructural protein NS5A. Gene. December 1996. Vol. 182, pages 203-211, see entire document.	1-3, 8-26
X	US 5,514,539 A (BUKH et al.) 07 May 1996, see entire document.	1-3, 8-26

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US97/09884

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☒ Claims Nos.: 4-7  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:  
  
The enclosed copy of claims 4, 5 were not legible and claims 6, 7 depend on those claims.
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.  
☐ No protest accompanied the payment of additional search fees.